

National Children's Study Research Day

A Semantically-Aware Approach to Data Comparability and Interoperability for the National Children's Study

Bethesda, MD
August 24, 2011

Agenda

- ▶ Metadata Specification and Maintenance in the NCS
- ▶ NCS Metadata Repository Purpose and Functionality
- ▶ Use and Future Directions

The National Children's Study requires metadata specification and maintenance to support the ongoing data lifecycle

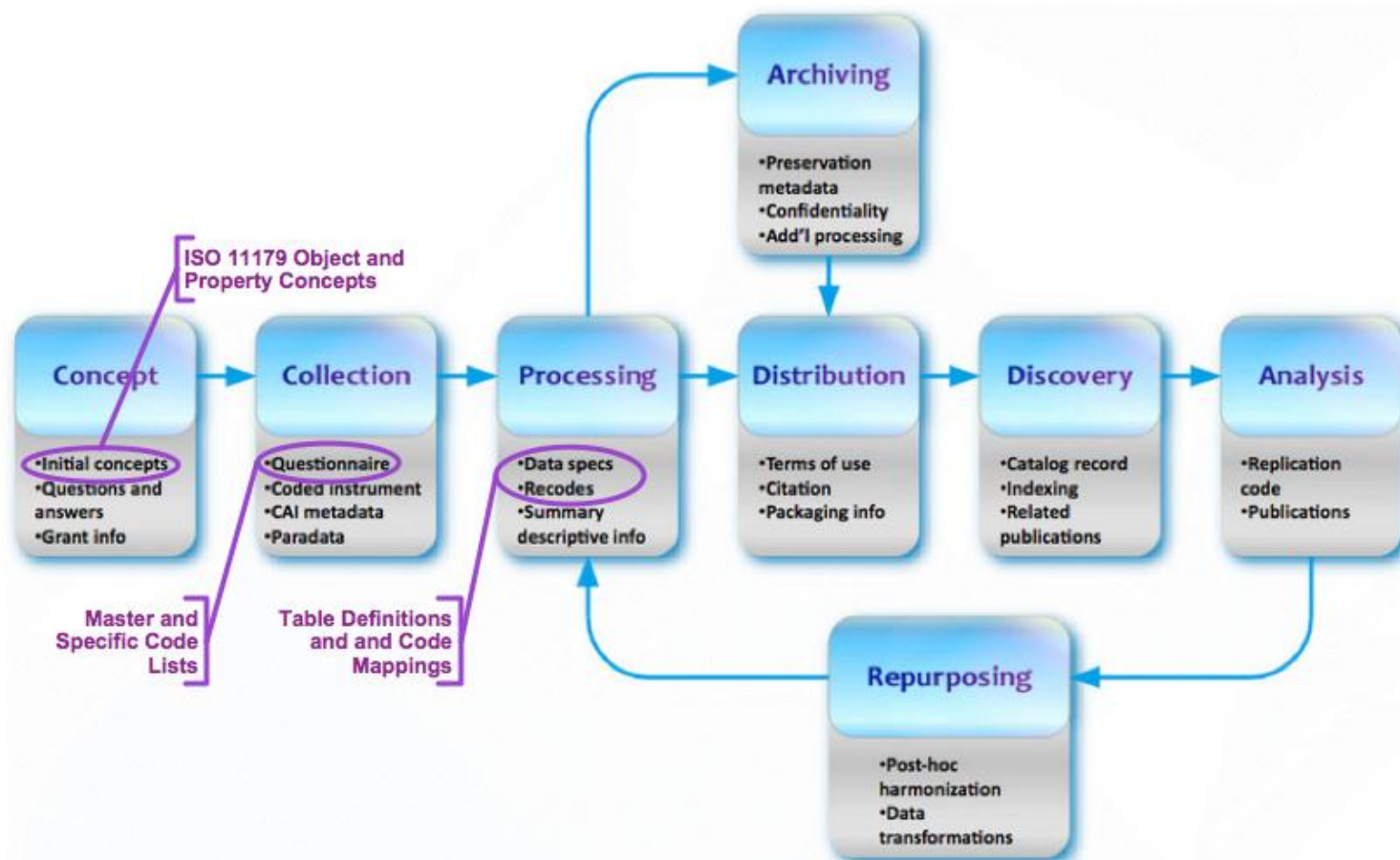
- ▶ For a longitudinal study such as the NCS, observations must be comparable over time and with other studies
 - Challenges to semantic interoperability or harmonization can occur due to incomparability or inconsistency of observations across successive versions of the same questionnaire and/or across observations in different questionnaires within the study
 - These challenges of internal harmonization can often be mitigated by post-harmonization (developing maps between value sets that guide recoding where necessary)
- ▶ To support this harmonization and mapping, an approach was established to building a Metadata Repository (MDR) for the study, to construct and maintain integrity of the Study metadata, provide proper versioning among data element specifications, and promote consistency in NCS workflows and informatics components that define, manage, and analyze the study data

The NCS has established a standards-based Metadata Repository in which definitions of data elements, value sets and forms are stored and versioned over time

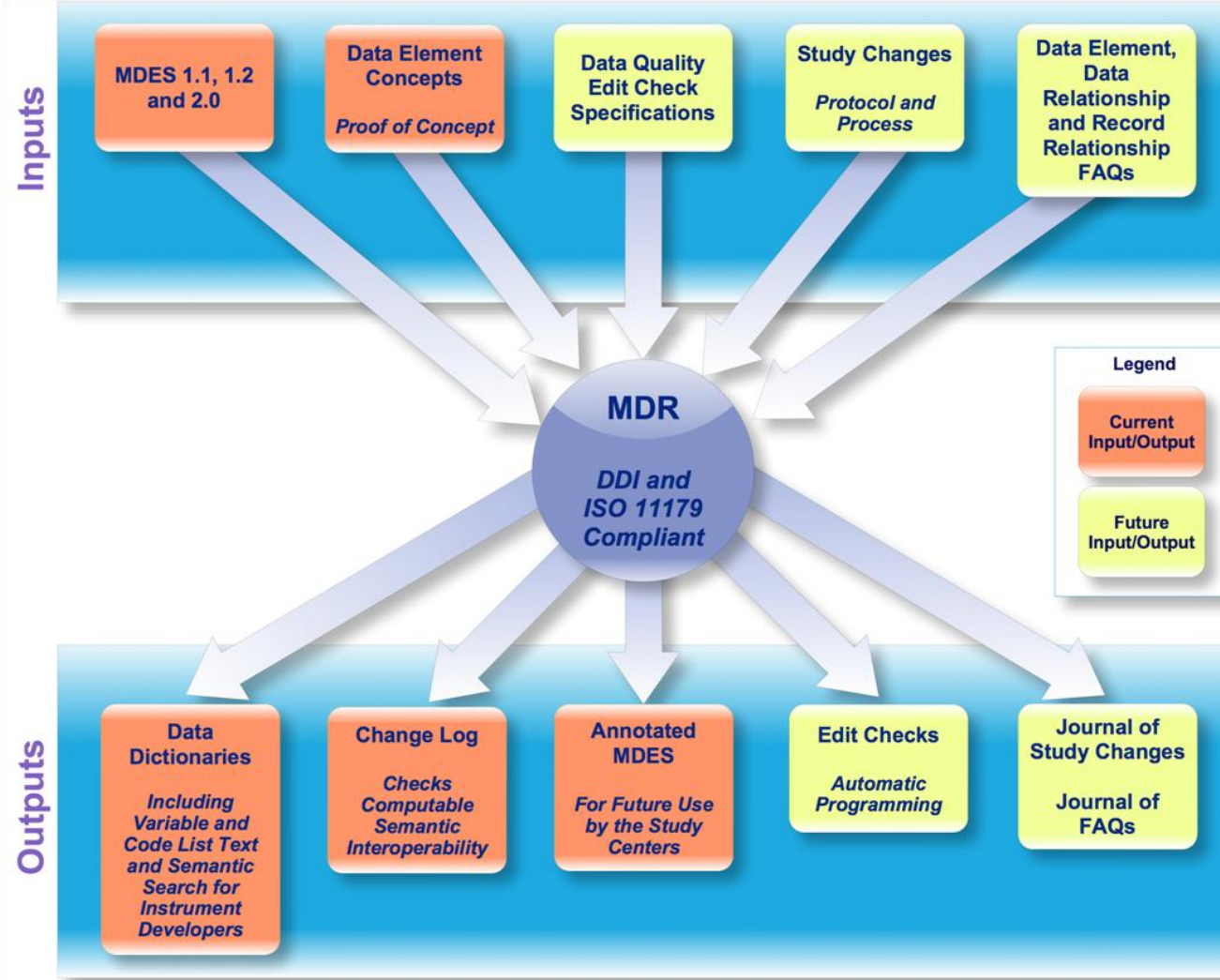
- ▶ The Master Data Elements Specification (MDES) is a data exchange specification for submitting operational and instrument data to a central Vanguard Data Repository (VDR)
- ▶ The MDR has taken the MDES as a starting point for extension, to support data life cycle phases after data collection and aggregation (for example, processing and analysis)
- ▶ The MDR extensions add scope to the MDES and conform to two international standards:
 - ISO 11179: Assures the interoperability of data elements within NCS and between NCS and other studies
 - Data Documentation Initiative (DDI) standard: Assures the traceability of data elements across the NCS data life cycle

The MDR takes the MDES as input and stores each version in a DDI compliant metadata database

- ▶ The MDR database has several storage areas in line with the DDI model shown here
- ▶ The MDR extensions tag data elements with concepts in the concept area
- ▶ The MDR extensions trace data element, table and code list definitions across successive versions of the MDES keeping track of them in the collection and processing areas



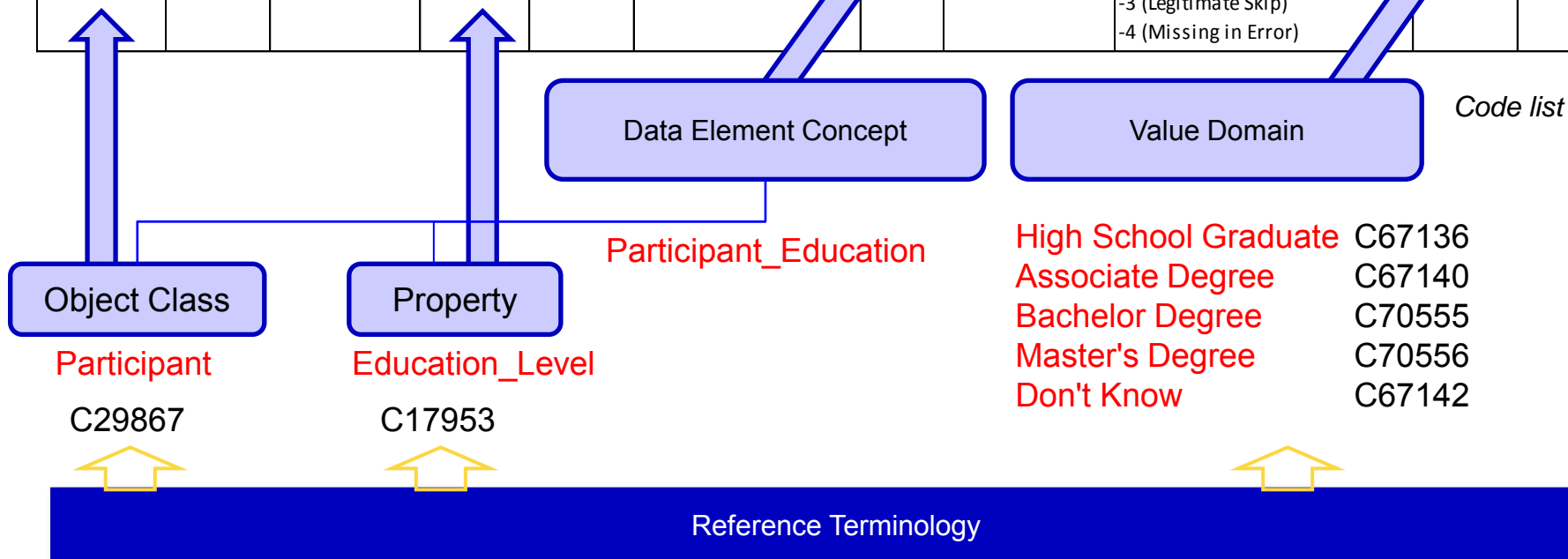
Various inputs to the MDR produce usable outputs to support study and operational analyses



A mapping of study variables in the MDR to reference terminology provides a linkage to commonly accepted standards

Master Data Elements Specification (MDES) Version 2.0
Education Level Variable Mapped to NCI Thesaurus Concepts

Object Class Concept	Concept Code	Concept Definition	Property Concept	Concept Code	Concept Definition	Variable Name	Variable Label/Definition	Format Constraint	Code List	Instrument Name
Participant	C29867	Someone who takes part in an activity.	Education Level	C17953	An indication of the years of schooling completed in graded public, private, or parochial schools, and in colleges, universities, or professional schools.	EDUC	Participant's level of education	1 (Less than High School) 2 (High School Diploma) 3 (Some college) 4 (Assoc Degree) 5 (Bach Degree) 6 (Post Grad Degree) -1 (Refused) -2 (Don't Know) -3 (Legitimate Skip) -4 (Missing in Error)	EDUCATION_LEVEL_C1	I9 - Pregnancy Visit 1 Interview (EH, PB, HI) - V1.0



The MDR produces a data dictionary that shows the linkages among NCS variables and reference concepts

National Children's Study Data Dictionary Version 2.0

Instrument Data Elements

4705 Data Elements
194 Table Definitions
415 Code Lists
10 Populations/Universes/Entities (Under Construction)
68 Concepts (Under Construction)

This site has four navigators. Select one:

[\[Data Elements\]](#)
[\[Code Lists / Value Sets\]](#)
[\[Populations/Universes/Entities\]](#)
[\[Concepts\]](#)

Concepts		
Name	Description	Concept ID
ASTHMA	A chronic respiratory disease manifested as difficulty breathing due to the narrowing of bronchial passageways.	C28397

NCI Thesaurus

ncit.nci.nih.gov/ncitbrowser/pages/concept_details.jsf?dictionary=NCIT

For quick access, place your bookmarks here on the bookmarks bar. [Import bookmarks now...](#) Other bookmarks

National Cancer Institute U.S. National Institutes of Health | www.cancer.gov

EVS Enterprise Vocabulary Services NCI Term Browser

Search ?

☒ Exact Match ☐ Begins With ☐ Contains
☒ Name/Code ☐ Property ☐ Relationship

Home | View Hierarchy | Subsets | Help Visited Concepts

Quick Links

Asthma (Code C28397)

[Suggest changes to this concept](#)

Terms & Properties Relationships Synonym Details View All

Terms and Properties

Definition: A chronic respiratory disease manifested as difficulty breathing due to the narrowing of bronchial passageways.

NCI-GLOSS Definition: A chronic disease in which the bronchial airways in the lungs become narrowed and swollen, making it difficult to breathe. Symptoms include wheezing, coughing, tightness in the chest, shortness of breath, and rapid breathing. An attack may be brought on by pet hair, dust, smoke, pollen, mold, exercise, cold air, or stress.

Preferred Name: Asthma

The data dictionary contains a change log that allows identification and reconciliation of inconsistencies in meaning across instrument versions

- ▶ To date, there are over 6,000 data elements in MDES v2.0
- ▶ Overall, there were 52 consequential changes with most of the problems occurring between Phase 1 and Phase 2 versions of the same data elements
- ▶ Recoding can mitigate most, though not all, of these issues

TWELVE_MTH_SAQ_FORMULA_BRAND_2															
This table was included in Version 1.2 as "TWELVE_MTH_SAQ_FORMULA_BRAND". Any changes to data elements in TWELVE_MTH_SAQ_FORMULA_BRAND_2 are listed below...															
Variable Name	Change														
	<p>NOTE: Previously the label at "47" was "Bright Beginnings NeoCare". Now the label is "Store brand Prebiotic like Target up & up".</p> <p>NOTE: "BABY_FORMULA_CL2" is not a superset of "BABY_FORMULA_CL1" The pair "48" and "Other type of formula" cannot be found in "BABY_FORMULA_CL2"</p> <p>NOTE: "BABY_FORMULA_CL2" is not a superset of "BABY_FORMULA_CL1" The pair "-3" and "Legitimate Skip" cannot be found in "BABY_FORMULA_CL2"</p>														
FORMULA_BRAND	<table> <tr> <th colspan="2">Computable Semantic Interoperability</th></tr> <tr> <th colspan="2">Code Map between "BABY_FORMULA_CL1" and "BABY_FORMULA_CL2"</th></tr> <tr> <th>BABY_FORMULA_CL1</th><th>BABY_FORMULA_CL2</th></tr> <tr> <td>40</td><td>1</td></tr> <tr> <td>41</td><td>2</td></tr> <tr> <td>42</td><td>3</td></tr> <tr> <td>43</td><td>4</td></tr> </table>	Computable Semantic Interoperability		Code Map between "BABY_FORMULA_CL1" and "BABY_FORMULA_CL2"		BABY_FORMULA_CL1	BABY_FORMULA_CL2	40	1	41	2	42	3	43	4
Computable Semantic Interoperability															
Code Map between "BABY_FORMULA_CL1" and "BABY_FORMULA_CL2"															
BABY_FORMULA_CL1	BABY_FORMULA_CL2														
40	1														
41	2														
42	3														
43	4														

Use and Future Directions

- ▶ The MDR provides a computable mechanism for maintaining study metadata for the NCS that can support versioning over time and track changes among data elements and code lists
- ▶ The data dictionary and its change log identify problems in meaning such as questions that are intended to measure the same concept whose value sets, however, are not immediately comparable
- ▶ Establishing consistency, versioning, and meaningful representation of the study metadata allows Study Centers to use local informatics while preserving the comparability and interoperability of study data
- ▶ The MDR supports the NCS as an integrated semantically-aware effort, allowing meaningful linkages of data and establishing connections to standards organizations and other studies

Acknowledgements

- ▶ Dr. Steven Hirschfeld, Associate Director for Clinical Research, NICHD
- ▶ Mr. David Songco, Chief Information Officer, NICHD

Booz Allen Hamilton:

- ▶ Sophia Kuan
- ▶ Marek Cyran
- ▶ Alexandra Shlionskaya
- ▶ Ipsita Das
- ▶ Ranjana Srivastava
- ▶ Michael Keller
- ▶ Daisy Guo

Thank You

- ▶ For more information please contact:

Daniela Smith

Booz Allen Hamilton

smith_daniela@bah.com

Jay Greenfield, PhD

Booz Allen Hamilton

greenfield_jay@bah.com